

# Network Utility Maximization over Partially Observable Markovian Channels

Chih-ping Li, *Student Member, IEEE* and Michael J. Neely, *Senior Member, IEEE*

**Abstract**—We consider a utility maximization problem over partially observable Markov ON/OFF channels. In this network instantaneous channel states are never known, and at most one user is selected for service in every slot according to the partial channel information provided by past observations. Solving the utility maximization problem directly is difficult because it involves solving partially observable Markov decision processes. Instead, we construct an approximate solution by optimizing the network utility only over a good constrained network capacity region rendered by stationary policies. Using a novel frame-based Lyapunov drift argument, we design a policy of admission control and user selection that stabilizes the network with utility that can be made arbitrarily close to the optimal in the constrained region. Equivalently, we are dealing with a high-dimensional restless bandit problem with a general functional objective over Markov ON/OFF restless bandits. Thus the network control algorithm developed in this paper serves as a new approximation methodology to attack such complex restless bandit problems.

## I. INTRODUCTION

This paper studies a multi-user wireless scheduling problem over partially observable environments. We consider a wireless uplink system serving  $N$  users via  $N$  independent Markov ON/OFF channels (see Fig. 1). Suppose time is slotted with

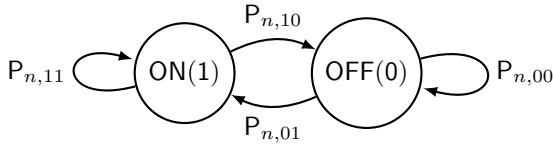


Fig. 1. The Markov ON/OFF chain for channel  $n \in \{1, 2, \dots, N\}$ .

normalized slots  $t \in \mathbb{Z}^+$ . Channel states are fixed in every slot, and can only change at slot boundaries. In every slot, the channel states are unknown, and at most one user is selected for transmission. The chosen user can successfully deliver a packet if the channel is ON, and zero otherwise. Since channels are ON/OFF, the state of the used channel is uncovered by an error-free ACK/NACK feedback at the end of the slot (failing to receive an ACK is regarded as a NACK). The states of each Markovian channel are correlated over time, and thus the revealed channel condition

Chih-ping Li (web: <http://www-scf.usc.edu/~chihpinl>) and Michael J. Neely (web: <http://www-rcf.usc.edu/~mjneely>) are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA.

This material is supported in part by one or more of the following: the DARPA IT-MANET program grant W911NF-07-0028, the NSF Career grant CCF-0747525, and continuing through participation in the Network Science Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory.

from ACK/NACK feedback provides partial information of future states, which can be used to improve user selection decisions and network performance. Our goal is to design a network control policy that maximizes a general network utility metric which is a function of the achieved throughput vector. Specifically, let  $y_n(t)$  be the amount of user- $n$  data served in slot  $t$ , and define the throughput  $\bar{y}_n$  for user  $n$  as  $\bar{y}_n \triangleq \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}[y_n(\tau)]$ . Let  $\Lambda$  be the *network capacity region* of the wireless uplink, defined as the closure of the set of all achievable throughput vectors  $\bar{\mathbf{y}} \triangleq (\bar{y}_n)_{n=1}^N$ . Then we seek to solve the following utility maximization problem:

$$\text{maximize: } g(\bar{\mathbf{y}}) \quad (1)$$

$$\text{subject to: } \bar{\mathbf{y}} \in \Lambda \quad (2)$$

where in the above we denote by  $g(\cdot)$  a generic utility function that is concave, continuous, nonnegative, and nondecreasing.

The problem (1)-(2) is very important to explore because it has many applications in various fields. In multi-user wireless scheduling, optimizing network utility over stochastic networks is first solved in [1], under the assumption that channel states are i.i.d. over slots and are known perfectly and instantly. The problem (1)-(2) we consider here generalizes the network utility maximization framework in [1] to networks with limiting channel probing capability (see [2], [3] and references therein) and delayed/uncertain channel state information (see [4]–[6] and references therein), in which we shall take advantage of channel memory [7] to improve network performance. In sequential decision making, (1)-(2) also captures an important class of restless bandit problems [8] in which each Markovian channel represents a two-state restless bandit, and packets served over a channel are rewards from playing the bandit. This class of Markov ON/OFF restless bandit problems has modern applications in opportunistic spectrum access in cognitive radio networks [9], [10] and target tracking of unmanned aerospace vehicles [11].

Solving the maximization problem (1)-(2) is difficult because  $\Lambda$  is unknown. In principle, we may compute  $\Lambda$  by locating its boundary points. However, they are solutions to  $N$ -dimensional Markov decision processes with information state vectors  $\omega(t) \triangleq (\omega_n(t))_{n=1}^N$ , where  $\omega_n(t)$  is the conditional probability that channel  $n$  is ON in slot  $t$  given the channel observation history. Namely, let  $s_n(t)$  denote the state of channel  $n$  in slot  $t$ . Then

$$\omega_n(t) \triangleq \Pr[s_n(t) = \text{ON} \mid \text{channel observation history}]. \quad (3)$$

We will show later  $\omega_n(t)$  takes values in a countably infinite set. Thus computing  $\Lambda$  and solving (1)-(2) seem to be infea-

sible.

Instead of solving (1)-(2), in this paper we adopt an achievable region approach to construct approximate solutions to (1)-(2). The key idea is two-fold. First, we explore the problem structure and construct an achievable throughput region  $\Lambda_{\text{int}} \subset \Lambda$  rendered by *good* stationary (possibly randomized) policies. Then we solve the constrained maximization problem:

$$\text{maximize: } g(\bar{\mathbf{y}}) \quad (4)$$

$$\text{subject to: } \bar{\mathbf{y}} \in \Lambda_{\text{int}} \quad (5)$$

as an approximation to (1)-(2). This approximation is practical because every throughput vector in  $\Lambda_{\text{int}}$  is attainable by simple stationary policies, and achieving feasible points outside  $\Lambda_{\text{int}}$  may require solving the much more complicated partially observable Markov decision processes (POMDPs) that relate to the original problem. Thus for the sake of simplicity and practicality, we shall regard  $\Lambda_{\text{int}}$  as our *operational* network capacity region.

Using the rich structure of the Markovian channels, in [12], [13] we have constructed a good achievable region  $\Lambda_{\text{int}}$  rendered by a special class of *randomized round robin* policies. It is important to note that we will maximize  $g(\bar{\mathbf{y}})$  only over this class of policies. Since every point in  $\Lambda_{\text{int}}$  can be achieved by one such policy (which we will show later), equivalently we are solving (4)-(5). We remark that solving (4)-(5) is decoupled from the construction of  $\Lambda_{\text{int}}$ . We will show in this paper that (4)-(5) can be solved. Therefore, the overall optimality of this achievable region approach depends on the proximity of the inner bound  $\Lambda_{\text{int}}$  to the full capacity region  $\Lambda$ .

The main contribution of this paper is that, using the Lyapunov optimization theory originally developed in [14], [15] and later generalized by [1], [16] for optimal stochastic control over wireless networks (see [17] for an introduction), we can solve (4)-(5) and develop optimal greedy algorithms. Specifically, using a novel Lyapunov drift argument, we construct a frame-based, queue-dependent network control algorithm of service allocation and admission control.<sup>1</sup> At the beginning of each frame, the admission controller decides how much new data to admit by solving a simple convex program.<sup>2</sup> The service allocation decision selects a randomized round robin policy by maximizing an *average MaxWeight* metric, and runs the policy for one round in the frame. We will show that this joint policy stabilizes the network and yields the achieved network utility  $g(\bar{\mathbf{y}})$  satisfying

$$g(\bar{\mathbf{y}}) \geq g(\bar{\mathbf{y}}^*) - \frac{B}{V_g}, \quad (6)$$

where  $g(\bar{\mathbf{y}}^*)$  is the optimal objective of (4)-(5),  $B > 0$  is a finite constant,  $V_g$  is a predefined positive control parameter, and we temporarily assume that all limits exist. By choosing  $V_g$  sufficiently large, we can approach the optimal utility  $g(\bar{\mathbf{y}}^*)$  arbitrarily well in (6), and thus solve (4)-(5).

Restless bandit problems with Markov ON/OFF bandits have been studied in [18]–[20], in which *index policies* [8],

<sup>1</sup> Admission control is used to facilitate the solution to the problem (4)-(5).

<sup>2</sup> The admission control decision decouples into  $N$  separable one-dimensional problems that are easily solved in real time in the case when  $g(\bar{\mathbf{y}})$  is a sum of one-dimensional utility functions for each user.

[21] are developed to maximize long-term average/discounted rewards. In this paper we extend this class of problems to having a general functional objective that needs to be maximized. This new problem is difficult to solve using existing approaches such as Whittle's index [8] or Markov decision theory [22], because they are typically limited to deal with problems with very simple objectives. The achievable region approach we develop in this paper solves (approximately) this extended problem, and thus could be viewed as a new approximation methodology to analyze similar complex restless bandit problems.

In the next section we introduce the detailed network model. Section III summarizes the construction of the inner bound  $\Lambda_{\text{int}}$  in [12], [13]. Our dynamic control algorithm is developed in Section IV, and the performance analysis is given in Section V.

## II. DETAILED NETWORK MODEL

In addition to the basic network model given in Section I, we suppose every channel  $n \in \{1, \dots, N\}$  evolves according to the transition probability matrix

$$\mathbf{P}_n = \begin{bmatrix} P_{n,00} & P_{n,01} \\ P_{n,10} & P_{n,11} \end{bmatrix},$$

where state ON is represented by 1 and OFF by 0, and  $P_{n,ij}$  denotes the transition probability from state  $i$  to  $j$ . We suppose every channel is *positively correlated over time*, so that an ON state is likely to be followed by another ON state. An equivalent mathematical definition is  $x_n \triangleq P_{n,01} + P_{n,10} < 1$  for all  $n$ . Let  $\mathbf{P}_n$  be known by both the network and user  $n$ .

We suppose every user has a data source of unlimited packets. In every slot, user  $n \in \{1, \dots, N\}$  admits  $r_n(t) \in [0, 1]$  packets from the source into a queue  $Q_n(t)$  of infinite capacity. For simplicity, we assume  $r_n(t)$  takes real values in  $[0, 1]$  for all  $n$ .<sup>3</sup> Define  $\mathbf{r}(t) \triangleq (r_n(t))_{n=1}^N$ . At the beginning of every slot, the network chooses and sends to the users one feasible admitted data vector  $\mathbf{r}(t)$  according to some admission policy. We let  $Q_n(t)$  and  $\mu_n(t) \in [0, 1]$  denote the queue backlog and the service rate of user  $n$  in slot  $t$ . Assume  $Q_n(0) = 0$  for all  $n$ . Then the queueing process  $\{Q_n(t)\}$  evolves as

$$Q_n(t+1) = \max[Q_n(t) - \mu_n(t), 0] + r_n(t). \quad (7)$$

The network keeps track of the backlog vector  $\mathbf{Q}(t) \triangleq (Q_n(t))_{n=1}^N$  in every slot. We say queue  $Q_n(t)$  is (strongly) stable if

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}[Q_n(\tau)] < \infty,$$

and the network is stable if all queues in the network are stable. Clearly a sufficient condition for stability is:

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{n=1}^N \mathbb{E}[Q_n(\tau)] < \infty. \quad (8)$$

Our goal is to design a policy that admits the right amount of packets into the network and serves them properly, so that

<sup>3</sup> We can accommodate the integer-value assumption of  $r_n(t)$  by introducing *auxiliary queues*; see [1] for an example.

the network is stable with utility that can be made arbitrarily close to the optimal solution to (4)-(5).

### III. A PERFORMANCE INNER BOUND

In this section we summarize the results in [12], [13] on constructing an achievable region  $\Lambda_{\text{int}}$  using randomized round robin policies. See [13] for detailed proofs.

#### A. Sufficient statistic

As discussed in [23, Chapter 5.4], the information state vector  $\omega(t)$  defined in (3) is a *sufficient statistic* of the network, meaning that it suffices to make optimal decisions based only on  $\omega(t)$  in every slot.

For channel  $n \in \{1, \dots, N\}$ , we denote by  $P_{n,ij}^{(k)}$  the  $k$ -step transition probability from state  $i$  to  $j$ , and  $\pi_{n,\text{ON}}$  its stationary probability of state ON. Since channels are positively correlated, we can show that  $P_{n,01}^{(k)}$  is nondecreasing and  $P_{n,11}^{(k)}$  is nonincreasing in  $k$ , and  $\pi_{n,\text{ON}} = \lim_{k \rightarrow \infty} P_{n,01}^{(k)} = \lim_{k \rightarrow \infty} P_{n,11}^{(k)}$ . For channel  $n$ , conditioning on the outcome of the last observation and when it was taken, it is easy to see that  $\omega_n(t)$  takes values in the countably infinite set  $\mathcal{W}_n \triangleq \{P_{n,01}^{(k)}, P_{n,11}^{(k)} : k \in \mathbb{N}\} \cup \{\pi_{n,\text{ON}}\}$ . Let  $n(t)$  be the channel observed in slot  $t$  via ACK/NACK feedback. The evolution of  $\omega_n(t)$  for each  $n$  then follows:

$$\omega_n(t+1) = \begin{cases} P_{n,01}, & \text{if } n = n(t), s_n(t) = \text{OFF} \\ P_{n,11}, & \text{if } n = n(t), s_n(t) = \text{ON} \\ \omega_n(t)P_{n,11} + (1 - \omega_n(t))P_{n,01}, & \text{if } n \neq n(t). \end{cases} \quad (9)$$

#### B. Randomized round robin

Let  $\Phi$  denote the set of all  $N$ -dimensional binary vectors excluding the zero vector  $\mathbf{0}$ . Every vector  $\phi \triangleq (\phi_n)_{n=1}^N \in \Phi$  stands for a collection of *active channels*, where we say channel  $n$  is active in  $\phi$  if  $\phi_n = 1$ . Let  $M(\phi)$  denote the number of 1's (or active channels) in  $\phi$ .

Consider the following *dynamic round robin* policy  $\text{RR}(\phi)$  that serves active channels in  $\phi$  possibly with different order in different rounds. This is the building block of the randomized round robin policies that we will introduce shortly.

##### Dynamic Round Robin Policy $\text{RR}(\phi)$ :

- 1) In each round, suppose an ordering of active channels in  $\phi$  is given.
- 2) When switching to active channel  $n$ , with probability  $P_{n,01}^{(M(\phi))}/\omega_n(t)$  keep transmitting packets over channel  $n$  until a NACK is received, and then switch to the next active channel. With probability  $1 - P_{n,01}^{(M(\phi))}/\omega_n(t)$ , transmit a dummy packet with no information content for one slot (used for channel sensing) and then switch to the next active channel.
- 3) Update  $\omega(t)$  according to (9) in every slot.

It is shown in [24] that, when channels have the same transition probability matrix, serving all channels by a greedy round robin policy maximizes the sum throughput of the network. Thus we shall get a good achievable throughput

region  $\Lambda_{\text{int}}$  by randomly mixing round robin policies, each of which serves a different subset of channels.

Consider the following randomized round robin that mixes  $\text{RR}(\phi)$  policies for different  $\phi$ :

##### Randomized Round Robin Policy RandRR:

- 1) Pick  $\phi \in \Phi \cup \{\mathbf{0}\}$  with probability  $\alpha_\phi$ , where  $\alpha_{\mathbf{0}} + \sum_{\phi \in \Phi} \alpha_\phi = 1$ .
- 2) If  $\phi \in \Phi$  is selected, run  $\text{RR}(\phi)$  for one round with the channel ordering of *least recently used first*. Then go to Step 1. If  $\phi = \mathbf{0}$ , idle the system for one slot and then go to Step 1.

For notational convenience, let  $\text{RR}(\mathbf{0})$  denote the operation of idling the system for one slot. For any  $\phi \in \Phi$ , we note that the  $\text{RR}(\phi)$  policy is feasible only if  $P_{n,01}^{(M(\phi))} \leq \omega_n(t)$  whenever we switch to active channel  $n$ . This condition is enforced in every RandRR policy by serving active channels in the order of *least recently used first* [13, Lemma 6]. Consequently, every RandRR is a feasible policy.<sup>4</sup> We note that the RandRR policies considered here are a superset of those in [13], because here we allow the additional idling operation. This enlarged policy space, however, has the same achievable throughput region as that in [13], because idle operations do not improve throughput. We generalize the RandRR policies here to ensure that every feasible point in  $\Lambda_{\text{int}}$  can be achieved by some RandRR policy. It is also helpful to note that, for any  $\phi \in \Phi$  and a fixed channel ordering, every  $\text{RR}(\phi)$  policy is a special case of the randomized round robin RandRR with  $\alpha_\phi = 1$  and 0 otherwise.

#### C. The achievable region

Next we summarize the achievable region rendered by randomized round robin policies.

**Theorem 1** ([12], [13]). *For each vector  $\phi \in \Phi$ , define the  $N$ -dimensional vector  $\eta^{(\phi)} \triangleq (\eta_n^{(\phi)})_{n=1}^N$  where*

$$\eta_n^{(\phi)} \triangleq \begin{cases} \frac{P_{n,01}(1 - (1 - x_n)^{M(\phi)})/(x_n P_{n,10})}{M(\phi) + \sum_{n: \phi_n=1} \frac{P_{n,01}(1 - (1 - x_n)^{M(\phi)})}{x_n P_{n,10}}}, & \text{if } \phi_n = 1 \\ 0, & \text{if } \phi_n = 0 \end{cases}$$

and  $x_n \triangleq P_{n,01} + P_{n,10}$ . Then the class of RandRR policies supports all throughput vectors  $\lambda$  in the set

$$\Lambda_{\text{int}} \triangleq \left\{ \lambda \mid \mathbf{0} \leq \lambda \leq \mu, \mu \in \text{conv} \left( \{ \eta^{(\phi)} \}_{\phi \in \Phi} \right) \right\},$$

where  $\text{conv}(A)$  denotes the convex hull of set  $A$ , and  $\leq$  is taken entrywise.

**Corollary 1.** *When channels have the same transition probability matrix so that  $\mathbf{P}_n = \mathbf{P}$  for all  $n$ , we have:*

$$\Lambda_{\text{int}} = \left\{ \lambda \mid \mathbf{0} \leq \lambda \leq \mu, \mu \in \text{conv} \left( \left\{ \frac{c_{M(\phi)}}{M(\phi)} \phi \right\}_{\phi \in \Phi} \right) \right\},$$

<sup>4</sup>The feasibility of RandRR policies is proved in [13] under the special case that there are no idle operations ( $\alpha_{\mathbf{0}} = 0$ ). Using the monotonicity of  $k$ -step transition probabilities  $\{P_{n,01}^{(k)}, P_{n,11}^{(k)}\}$ , the feasibility can be similarly proved for the generalized RandRR policies considered here.

where

$$c_{M(\phi)} \triangleq \frac{P_{01}(1 - (1 - x)^{M(\phi)})}{x P_{10} + P_{01}(1 - (1 - x)^{M(\phi)})}, \quad x = P_{01} + P_{10}, \quad (10)$$

and we have dropped the subscript  $n$  due to channel symmetry.

The closeness of the inner bound  $\Lambda_{\text{int}}$  and the full capacity region  $\Lambda$  is quantified in [13] in the special case that channels have the same transition probability matrix. For any feasible direction  $\mathbf{v}$ , it can be shown that as  $\mathbf{v}$  becomes more symmetric, or forms a smaller angle with the 45-degree line, the loss of the sum throughput of the inner boundary point in direction  $\mathbf{v}$  decreases to zero geometrically fast, provided that the network serves a large number of users.

Next, that RandRR policies considered in this paper are random mixings of those in [13] and idle operations leads to the next corollary.

**Corollary 2.** *Every throughput vector in  $\Lambda_{\text{int}}$  can be achieved by some RandRR policy.*

#### D. A two-user example

Consider a two-user system with symmetric channels with  $P_{01} = P_{10} = 0.2$ . From Corollary 1

$$\Lambda_{\text{int}} = \left\{ \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \left| \begin{array}{l} 0 \leq \lambda_n \leq \mu_n, \text{ for } 1 \leq n \leq 2, \\ \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \in \text{conv} \left( \left\{ \begin{bmatrix} c_2/2 \\ c_2/2 \end{bmatrix}, \begin{bmatrix} c_1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ c_1 \end{bmatrix} \right\} \right) \right. \right\}.$$

where  $c_1$  and  $c_2$  are defined in (10). Fig. 2 shows the closeness of  $\Lambda_{\text{int}}$  and  $\Lambda$  in this example. We note that points  $B$ ,

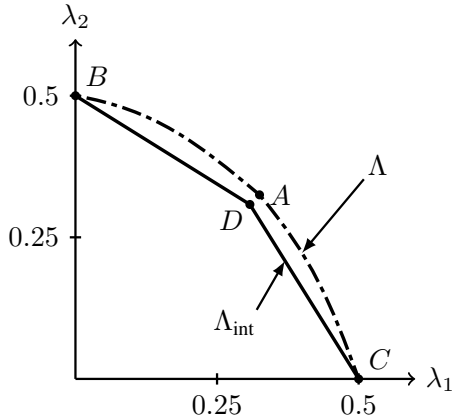


Fig. 2. The closeness of  $\Lambda_{\text{int}}$  and  $\Lambda$ .

$A$ , and  $C$  maximize the sum throughput of the network in directions  $(0, 1)$ ,  $(1, 1)$ , and  $(1, 0)$ , respectively [24]. Therefore the boundary of the (unknown) full capacity region  $\Lambda$  is a concave curve connecting these points.

#### IV. NETWORK UTILITY MAXIMIZATION

From Theorem 1, the constrained problem (4)-(5) is a well-defined convex program. However, solving (4)-(5) remains difficult because the representation of  $\Lambda_{\text{int}}$  via a convex hull of  $(2^N - 1)$  throughput vectors is very complicated. Next we

solve (4)-(5) by admission control and service allocation in the network. We will use the Lyapunov optimization theory to construct a dynamic policy that *learns* a near-optimal solution to (4)-(5), where the closeness to the true optimality is controlled by a positive control parameter  $V_g$ .

##### A. Constructing Lyapunov drift

We start with constructing a frame-based Lyapunov *drift-minus-utility* inequality over a frame of size  $T$ , where  $T$  is possibly random but has a finite second moment bounded by a constant  $C$  so that  $C \geq \mathbb{E}[T^2 | \mathbf{Q}(t)]$  for all  $t$  and all possible  $\mathbf{Q}(t)$ . Define  $B \triangleq NC$ . The result will shed light on the structure of our desired policy. By iteratively applying (7), it is not hard to show that

$$Q_n(t+T) \leq \max \left[ Q_n(t) - \sum_{\tau=0}^{T-1} \mu_n(t+\tau), 0 \right] + \sum_{\tau=0}^{T-1} r_n(t+\tau) \quad (11)$$

for each  $n \in \{1, \dots, N\}$ . We define the *Lyapunov function*

$$L(\mathbf{Q}(t)) \triangleq \frac{1}{2} \sum_{n=1}^N Q_n^2(t)$$

and the *T-slot Lyapunov drift*

$$\Delta_T(\mathbf{Q}(t)) \triangleq \mathbb{E}[L(\mathbf{Q}(t+T)) - L(\mathbf{Q}(t)) | \mathbf{Q}(t)],$$

where the expectation is over the randomness of the network in the frame, including that of  $T$ . By taking the following steps: (1) take square of (11) for each  $n$ ; (2) use the inequalities

$$\begin{aligned} \max[a - b, 0] &\leq a, \quad \forall a \geq 0, \\ (\max[a - b, 0])^2 &\leq (a - b)^2, \quad \mu_n(t) \leq 1, \quad r_n(t) \leq 1, \end{aligned}$$

to simplify terms; (3) sum all resulting inequalities; (4) take conditional expectation on  $\mathbf{Q}(t)$ , we can show

$$\begin{aligned} \Delta_T(\mathbf{Q}(t)) &\leq B \\ &- \mathbb{E} \left[ \sum_{n=1}^N Q_n(t) \left[ \sum_{\tau=0}^{T-1} \mu_n(t+\tau) - r_n(t+\tau) \right] | \mathbf{Q}(t) \right]. \end{aligned} \quad (12)$$

By subtracting from both sides of (12) the weighted sum utility

$$V_g \mathbb{E} \left[ \sum_{\tau=0}^{T-1} g(\mathbf{r}(t+\tau)) | \mathbf{Q}(t) \right],$$

where  $V_g > 0$  is a predefined control parameter, we get

$$\begin{aligned} \Delta_T(\mathbf{Q}(t)) - V_g \mathbb{E} \left[ \sum_{\tau=0}^{T-1} g(\mathbf{r}(t+\tau)) | \mathbf{Q}(t) \right] \\ \leq B - \sum_{n=1}^N Q_n(t) \mathbb{E} \left[ \sum_{\tau=0}^{T-1} \mu_n(t+\tau) | \mathbf{Q}(t) \right] \\ - \mathbb{E} \left[ \sum_{\tau=0}^{T-1} \left[ V_g g(\mathbf{r}(t+\tau)) - \sum_{n=1}^N Q_n(t) r_n(t+\tau) \right] | \mathbf{Q}(t) \right]. \end{aligned} \quad (13)$$

The above inequality gives an upper bound on the *drift-minus-utility* expression at the left side of (13), and holds for any

scheduling policy over a frame of any size  $T$ .

### B. Network control policy

Let  $f(\mathbf{Q}(t))$  and  $g(\mathbf{Q}(t))$  denote the second-to-last and the last term of (13):

$$\begin{aligned} f(\mathbf{Q}(t)) &\triangleq \sum_{n=1}^N Q_n(t) \mathbb{E} \left[ \sum_{\tau=0}^{T-1} \mu_n(t+\tau) \mid \mathbf{Q}(t) \right] \\ g(\mathbf{Q}(t)) &\triangleq \mathbb{E} \left[ \sum_{\tau=0}^{T-1} \left[ V_g g(\mathbf{r}(t+\tau)) \right. \right. \\ &\quad \left. \left. - \sum_{n=1}^N Q_n(t) r_n(t+\tau) \right] \mid \mathbf{Q}(t) \right], \end{aligned}$$

and (13) is equivalent to

$$\begin{aligned} \Delta_T(\mathbf{Q}(t)) - V_g \mathbb{E} \left[ \sum_{\tau=0}^{T-1} g(\mathbf{r}(t+\tau)) \mid \mathbf{Q}(t) \right] \\ \leq B - f(\mathbf{Q}(t)) - g(\mathbf{Q}(t)). \end{aligned} \quad (14)$$

After observing the current backlog vector  $\mathbf{Q}(t)$ , we seek to maximize over all feasible policies the average

$$\frac{f(\mathbf{Q}(t)) + g(\mathbf{Q}(t))}{\mathbb{E}[T \mid \mathbf{Q}(t)]} \quad (15)$$

over a frame of size  $T$ . Every feasible policy here consists of: (1) an admission policy that admits  $r_n(t+\tau)$  packets to user  $n$  in every slot of the frame, and (2) a randomized round robin RandRR policy introduced in Section III-B that serves a set of active users and decides the service rates  $\mu_n(t+\tau)$  in the frame. The random frame size  $T$  in (15) is the length of one transmission round under the candidate RandRR policy, and its distribution depends on the backlog vector  $\mathbf{Q}(t)$  via the queue-dependent choice of RandRR. We will show later that the novel performance metric (15) helps to achieve near-optimal network utility.

We simplify the procedure of maximizing (15) in the following, and the result is our network control algorithm. In  $g(\mathbf{Q}(t))$ , we observe that the optimal choices of the admitted data vectors  $\mathbf{r}(t+\tau)$  are independent of both the frame size  $T$  and the rate allocations  $\mu_n(t+\tau)$  in  $f(\mathbf{Q}(t))$ . Thus,  $\mathbf{r}(t+\tau)$  can be optimized separately. Specifically, the optimal values of  $\mathbf{r}(t+\tau)$  shall be the same for all  $\tau \in \{0, \dots, T-1\}$  and are the solution to

$$\text{maximize: } V_g g(\mathbf{r}(t)) - \sum_{n=1}^N Q_n(t) r_n(t) \quad (16)$$

$$\text{subject to: } r_n(t) \in [0, 1], \quad \forall n \in \{1, \dots, N\} \quad (17)$$

which only depends on the backlog vector  $\mathbf{Q}(t)$  at the beginning of the current frame and the predefined control parameter  $V_g$ . We note that if  $g(\cdot)$  is a sum of individual utilities so that  $g(\mathbf{r}(t)) = \sum_{n=1}^N g_n(r_n(t))$ , (16)-(17) decouples into  $N$  one-dimensional convex programs, each of which maximizes  $V_g g_n(r_n(t)) - Q_n(t) r_n(t)$  over  $r_n(t) \in [0, 1]$ , which can be solved efficiently in real time. Let  $h^*(\mathbf{Q}(t))$  be the resulting optimal objective of (16)-(17). It follows that

$$g(\mathbf{Q}(t)) = \mathbb{E}[T \mid \mathbf{Q}(t)] h^*(\mathbf{Q}(t))$$

and (15) is equal to

$$\frac{f(\mathbf{Q}(t))}{\mathbb{E}[T \mid \mathbf{Q}(t)]} + h^*(\mathbf{Q}(t)). \quad (18)$$

The sum (18) indicates that finding the optimal admission policy is independent of finding the optimal randomized round robin policy. It remains to maximize the first term of (18) over all RandRR policies.

Next we evaluate the first term of (18) under a fixed RandRR policy with parameters  $\{\alpha_\phi\}_{\phi \in \Phi \cup \{0\}}$ . Conditioning on the choice of  $\phi$ , we get

$$f(\mathbf{Q}(t)) = \sum_{\phi \in \Phi \cup \{0\}} \alpha_\phi f(\mathbf{Q}(t), \text{RR}(\phi)),$$

where  $f(\mathbf{Q}(t), \text{RR}(\phi))$  denotes the term  $f(\mathbf{Q}(t))$  evaluated under policy  $\text{RR}(\phi)$  (recall that  $\text{RR}(\phi)$  is a special case of the RandRR policy). Similarly, by conditioning we can show

$$\mathbb{E}[T] = \mathbb{E}[T \mid \mathbf{Q}(t)] = \sum_{\phi \in \Phi \cup \{0\}} \alpha_\phi \mathbb{E}[T_{\text{RR}(\phi)}],^5$$

where  $T_{\text{RR}(\phi)}$  denotes the duration of one transmission round under the  $\text{RR}(\phi)$  policy. It follows that

$$\frac{f(\mathbf{Q}(t))}{\mathbb{E}[T \mid \mathbf{Q}(t)]} = \frac{\sum_{\phi \in \Phi \cup \{0\}} \alpha_\phi f(\mathbf{Q}(t), \text{RR}(\phi))}{\sum_{\phi \in \Phi \cup \{0\}} \alpha_\phi \mathbb{E}[T_{\text{RR}(\phi)}]}. \quad (19)$$

The next lemma shows there always exists a  $\text{RR}(\phi)$  policy maximizing (19) over all RandRR policies. Therefore it suffices to focus only on  $\text{RR}(\phi)$  policies.

**Lemma 1.** *We index  $\text{RR}(\phi)$  policies for all  $\phi \in \Phi \cup \{0\}$ . For the  $\text{RR}(\phi)$  policy with index  $k$ , define*

$$f_k \triangleq f(\mathbf{Q}(t), \text{RR}(\phi)), \quad D_k \triangleq \mathbb{E}[T_{\text{RR}(\phi)}].$$

*Without loss of generality, assume*

$$\frac{f_1}{D_1} \geq \frac{f_k}{D_k}, \quad \forall k \in \{2, 3, \dots, 2^N\}.$$

*Then for any probability distribution  $\{\alpha_k\}_{k \in \{1, \dots, 2^N\}}$  with  $\alpha_k \geq 0$  and  $\sum_k \alpha_k = 1$ , we have*

$$\frac{f_1}{D_1} \geq \frac{\sum_{k=1}^{2^N} \alpha_k f_k}{\sum_{k=1}^{2^N} \alpha_k D_k}.$$

*Proof of Lemma 1:* Fact 1: Let  $\{a_1, a_2, b_1, b_2\}$  be four positive numbers, and suppose there is a bound  $z$  such that  $a_1/b_1 \leq z$  and  $a_2/b_2 \leq z$ . Then for any probability  $\theta$  (where  $0 \leq \theta \leq 1$ ), we have:

$$\frac{\theta a_1 + (1-\theta)a_2}{\theta b_1 + (1-\theta)b_2} \leq z. \quad (20)$$

We prove Lemma 1 by induction and (20). Initially, for any  $\alpha_1, \alpha_2 \geq 0$ ,  $\alpha_1 + \alpha_2 = 1$ , from  $f_1/D_1 \geq f_2/D_2$  we get

$$\frac{f_1}{D_1} \geq \frac{\alpha_1 f_1 + \alpha_2 f_2}{\alpha_1 D_1 + \alpha_2 D_2}.$$

<sup>5</sup>Given a fixed policy RandRR, the frame size  $T$  no longer depends on the backlog vector  $\mathbf{Q}(t)$ . Therefore  $\mathbb{E}[T] = \mathbb{E}[T \mid \mathbf{Q}(t)]$ .

For some  $K > 2$ , assume

$$\frac{f_1}{D_1} \geq \frac{\sum_{k=1}^{K-1} \alpha_k f_k}{\sum_{k=1}^{K-1} \alpha_k D_k} \quad (21)$$

holds for any probability distribution  $\{\alpha_k\}_{k=1}^{K-1}$ . It follows that, for any probability distribution  $\{\alpha_k\}_{k=1}^K$ , we get

$$\frac{\sum_{k=1}^K \alpha_k f_k}{\sum_{k=1}^K \alpha_k D_k} = \frac{(1 - \alpha_K) \left[ \sum_{k=1}^{K-1} \frac{\alpha_k}{1 - \alpha_K} f_k \right] + \alpha_K f_K}{(1 - \alpha_K) \left[ \sum_{k=1}^{K-1} \frac{\alpha_k}{1 - \alpha_K} D_k \right] + \alpha_K D_K} \stackrel{(a)}{\leq} \frac{f_1}{D_1}$$

where (a) is from Fact 1, noting that  $f_1/D_1 \geq f_K/D_K$  and

$$\frac{f_1}{D_1} \geq \frac{\sum_{k=1}^{K-1} \frac{\alpha_k}{1 - \alpha_K} f_k}{\sum_{k=1}^{K-1} \frac{\alpha_k}{1 - \alpha_K} D_k},$$

where the above holds by the induction assumption (21). ■

From Lemma 1, next we evaluate  $f(\mathbf{Q}(t))/\mathbb{E}[T | \mathbf{Q}(t)]$  for a given  $\text{RR}(\phi)$  policy. Again we have  $\mathbb{E}[T | \mathbf{Q}(t)] = \mathbb{E}[T]$ .

In the special case  $\phi = \mathbf{0}$ , we get  $f(\mathbf{Q}(t))/\mathbb{E}[T | \mathbf{Q}(t)] = 0$ . Otherwise, fix some  $\phi \in \Phi$ . For each active channel  $n$  in  $\phi$ , we denote by  $L_n^\phi$  the amount of time the network stays with user  $n$  in one round of  $\text{RR}(\phi)$ . It is shown in [13, Corollary 1] that  $L_n^\phi$  has the probability distribution

$$L_n^\phi = \begin{cases} 1 & \text{with prob. } 1 - P_{n,01}^{(M(\phi))} \\ j \geq 2 & \text{with prob. } P_{n,01}^{(M(\phi))} (P_{n,11})^{(j-2)} P_{n,10}, \end{cases} \quad (22)$$

and

$$\mathbb{E}[L_n^\phi] = 1 + \frac{P_{n,01}^{(M(\phi))}}{P_{n,10}}. \quad (23)$$

It follows that under the  $\text{RR}(\phi)$  policy we have

$$\begin{aligned} \mathbb{E}[T] &= \mathbb{E}[T | \mathbf{Q}(t)] = \sum_{n: \phi_n = 1} \mathbb{E}[L_n^\phi], \\ \mathbb{E}\left[\sum_{\tau=0}^{T-1} \mu_n(t + \tau) | \mathbf{Q}(t)\right] &= \begin{cases} \mathbb{E}[L_n^\phi] - 1 & \text{if } \phi_n = 1 \\ 0 & \text{if } \phi_n = 0 \end{cases} \end{aligned}$$

and thus

$$\frac{f(\mathbf{Q}(t))}{\mathbb{E}[T | \mathbf{Q}(t)]} = \frac{\sum_{n=1}^N Q_n(t) \mathbb{E}[L_n^\phi - 1] \phi_n}{\sum_{n=1}^N \mathbb{E}[L_n^\phi] \phi_n}. \quad (24)$$

The above simplifications lead to the next network control algorithm that maximizes (15) in a frame-by-frame basis over all feasible admission and randomized round robin policies.

#### Queue-dependent Round Robin for Network Utility Maximization (QRRNUM):

- 1) At the beginning of a transmission round, observe the current backlog vector  $\mathbf{Q}(t)$  and solve the convex program (16)-(17). Let  $\mathbf{r}^{\text{QRR}}(t) \triangleq (r_n^{\text{QRR}}(t))_{n=1}^N$  be the optimal solution.
- 2) Let  $\phi^{\text{QRR}}(t)$  be the maximizer of (24) over all  $\phi \in \Phi$ . If the resulting optimal objective is larger than zero, execute policy  $\text{RR}(\phi^{\text{QRR}}(t))$  for one round, with the channel ordering of least recently used first. Otherwise, idle the system for one slot. At the same time, admit  $r_n^{\text{QRR}}(t)$  packets to user  $n$  in every slot of the current round. At the end of the round, go to Step 1).

The most complex part of the QRRNUM algorithm is to maximize (24) in Step 2, where in general all  $(2^N - 1)$  choices of vector  $\phi \in \Phi$  need to be examined, resulting in exponential complexity. In the special case that channels have the same transition probability matrix, the QRRNUM algorithm reduces to a polynomial time policy, and the following steps find the maximizer  $\phi^{\text{QRR}}(t)$  of (24):

- 1) Re-index  $Q_n(t)$  so that  $Q_1(t) \geq Q_2(t) \geq \dots \geq Q_N(t)$ .
- 2) For each  $K \in \{1, \dots, N\}$ , compute

$$\frac{P_{01}^{(K)}}{K(P_{10} + P_{01}^{(K)})} \sum_{k=1}^K Q_k(t) \quad (25)$$

and let  $K^{\text{QRR}}$  be the maximizer of (25) over  $K$ .

- 3) If  $\mathbf{Q}(t)$  is the zero vector, let  $\phi^{\text{QRR}}(t) = \mathbf{0}$ . Otherwise, let  $\phi^{\text{QRR}}(t)$  be the binary vector with the first  $K^{\text{QRR}}$  components being 1 and 0 otherwise.

Another way to have an efficient QRRNUM algorithm, especially when  $N$  is large, is to restrict to a subset of  $\text{RR}(\phi)$  policies. For example, consider those in every transmission round only serve 2 or 0 users. Although the associated new achievable region  $\Lambda_{\text{int}}$  (can be found as a corollary of Theorem 1) will be smaller, the resulting QRRNUM algorithm has polynomial time complexity because we only need to consider  $N(N-1)/2$  choices of  $\phi$  in every round.

#### V. PERFORMANCE ANALYSIS

In the QRRNUM policy, let  $t_{k-1}$  and  $T_k$  be the beginning and the duration of the  $k$ th transmission round. We have  $T_k = t_k - t_{k-1}$  and  $t_k = \sum_{i=1}^k T_i$  for all  $k \in \mathbb{N}$ . Assume  $t_0 = 0$ . Every  $T_k$  is the length of a transmission round of some  $\text{RR}(\phi)$  policy. Define  $T_{\max}$  as the length of a transmission round of the policy  $\text{RR}(1)$  that serves all channels in every round. Then for each  $k \in \mathbb{N}$ , we can show that  $T_{\max}$  and  $T_{\max}^2$  is stochastically larger than  $T_k$  and  $T_k^2$ , respectively.<sup>6</sup> As a result, we have

$$\mathbb{E}[T_k] \leq \mathbb{E}[T_{\max}] < \infty, \quad \mathbb{E}[T_k^2] \leq \mathbb{E}[T_{\max}^2] < \infty. \quad (26)$$

The next theorem shows the performance of the QRRNUM algorithm.

**Theorem 2.** Let  $\mathbf{y}(t) = (y_n(t))_{n=1}^N$  be the vector of served packets for each user in slot  $t$ ;  $y_n(t) = \min[Q_n(t), \mu_n(t)]$ . Define constant  $B \triangleq N \mathbb{E}[T_{\max}^2]$ . Then for any given positive control parameter  $V_g > 0$ , the QRRNUM algorithm stabilizes the network and yields average network utility satisfying

$$\liminf_{t \rightarrow \infty} g\left(\frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}[\mathbf{y}(\tau)]\right) \geq g(\bar{\mathbf{y}}^*) - \frac{B}{V_g}, \quad (27)$$

where  $g(\bar{\mathbf{y}}^*)$  is the optimal network utility and the solution to the constrained restless bandit problem (4)-(5). By taking  $V_g$  sufficiently large, the QRRNUM algorithm achieves network utility arbitrarily close to the optimal  $g(\bar{\mathbf{y}}^*)$ , and thus solves (4)-(5).

<sup>6</sup>If  $\text{RR}(\phi)$  for some  $\phi \in \Phi$  is used in  $T_k$ , the stochastic ordering between  $T_{\max}$  and  $T_k$  can be shown by noting that  $T_k = \sum_{n: \phi_n = 1} L_n^\phi$ , where  $L_n^\phi$  is defined in (22). Otherwise, we have  $\phi = \mathbf{0}$  and  $T_k = 1 \leq T_{\max}$ .

*Proof of Theorem 2:* Analyzing the performance of the QRRNUM algorithm relies on comparing it to a near-optimal feasible solution. We will adopt the approach in [1] but generalize it to a frame-based analysis.

For some  $\epsilon > 0$ , consider the  $\epsilon$ -constrained version of (4)-(5):

$$\text{maximize: } g(\bar{\mathbf{y}}) \quad (28)$$

$$\text{subject to: } \bar{\mathbf{y}} \in \Lambda_{\text{int}}(\epsilon) \quad (29)$$

where  $\Lambda_{\text{int}}(\epsilon)$  is the achievable region  $\Lambda_{\text{int}}$  stripping an “ $\epsilon$ -layer” off the boundary:

$$\Lambda_{\text{int}}(\epsilon) \triangleq \{\bar{\mathbf{y}} \mid \bar{\mathbf{y}} + \epsilon \mathbf{1} \in \Lambda_{\text{int}}\},$$

where  $\mathbf{1}$  is an all-one vector. Notice that  $\Lambda_{\text{int}}(\epsilon) \rightarrow \Lambda_{\text{int}}$  as  $\epsilon \rightarrow 0$ . Let  $\bar{\mathbf{y}}^*(\epsilon) = (\bar{y}_n^*(\epsilon))_{n=1}^N$  and  $\bar{\mathbf{y}}^* = (\bar{y}_n^*)_{n=1}^N$  be the optimal solution to the  $\epsilon$ -constrained problem (28)-(29) and the constrained restless bandit problem (4)-(5), respectively. For simplicity, we assume  $\bar{\mathbf{y}}_\epsilon^* \rightarrow \bar{\mathbf{y}}^*$  as  $\epsilon \rightarrow 0$ .<sup>7</sup>

From Corollary 2, there exists a randomized round robin that yields the throughput vector  $\bar{\mathbf{y}}_\epsilon^* + \epsilon \mathbf{1}$  (note that  $\bar{\mathbf{y}}_\epsilon^* + \epsilon \mathbf{1} \in \Lambda_{\text{int}}$ ), and we denote this policy by  $\text{RandRR}_\epsilon^*$ . Let  $T_\epsilon^*$  denotes the length of one transmission round under  $\text{RandRR}_\epsilon^*$ . Then we have for each  $n \in \{1, \dots, N\}$

$$\mathbb{E} \left[ \sum_{\tau=0}^{T_\epsilon^*-1} \mu_n(t+\tau) \mid \mathbf{Q}(t) \right] \geq (\bar{y}_n^*(\epsilon) + \epsilon) \mathbb{E} [T_\epsilon^*] \quad (30)$$

from renewal reward theory. That is, we may consider a renewal reward process where renewal epochs are time instants at which  $\text{RandRR}_\epsilon^*$  starts a new round of transmission (with renewal period  $T_\epsilon^*$ ), and rewards are the allocated service rates  $\mu_n(t+\tau)$ .<sup>8</sup> Then the average service rate is simply the average sum reward over a renewal period divided by the average renewal duration  $\mathbb{E} [T_\epsilon^*]$ . Then (30) holds because this average service rate is greater than or equal to  $(\bar{y}_n^*(\epsilon) + \epsilon)$ .

Combining  $\text{RandRR}_\epsilon^*$  with the admission policy  $\sigma^*$  that sets  $r_n(t+\tau) = \bar{y}_n^*(\epsilon)$  for all  $n$  and  $\tau \in \{0, \dots, T_\epsilon^* - 1\}$ ,<sup>9</sup> we get

$$f_\epsilon^*(\mathbf{Q}(t)) \geq \mathbb{E} [T_\epsilon^*] \sum_{n=1}^N Q_n(t) (\bar{y}_n^*(\epsilon) + \epsilon) \quad (31)$$

$$g_\epsilon^*(\mathbf{Q}(t)) = \mathbb{E} [T_\epsilon^*] \left[ V_g g(\bar{\mathbf{y}}^*(\epsilon)) - \sum_{n=1}^N Q_n(t) \bar{y}_n^*(\epsilon) \right] \quad (32)$$

where (31)(32) are  $f(\mathbf{Q}(t))$  and  $g(\mathbf{Q}(t))$  evaluated under  $\text{RandRR}_\epsilon^*$  and  $\sigma^*$ , respectively.

Since the QRRNUM policy maximizes (15), evaluating (15)

under both QRRNUM and the policy  $(\text{RandRR}_\epsilon^*, \sigma^*)$  yields

$$\begin{aligned} & f_{\text{QRRNUM}}(\mathbf{Q}(t_k)) + g_{\text{QRRNUM}}(\mathbf{Q}(t_k)) \\ & \geq \mathbb{E} [T_{k+1} \mid \mathbf{Q}(t_k)] \frac{f_\epsilon^*(\mathbf{Q}(t_k)) + g_\epsilon^*(\mathbf{Q}(t_k))}{\mathbb{E} [T_\epsilon^*]} \\ & \stackrel{(a)}{\geq} \mathbb{E} [T_{k+1} \mid \mathbf{Q}(t_k)] \left[ V_g g(\bar{\mathbf{y}}^*(\epsilon)) + \epsilon \sum_{n=1}^N Q_n(t_k) \right] \\ & = \mathbb{E} \left[ T_{k+1} \left( V_g g(\bar{\mathbf{y}}^*(\epsilon)) + \epsilon \sum_{n=1}^N Q_n(t_k) \right) \mid \mathbf{Q}(t_k) \right], \end{aligned} \quad (33)$$

where (a) is from (31)(32). The drift-minus-utility bound (14) under the QRRNUM policy in the  $(k+1)$ th round of transmission yields

$$\begin{aligned} \Delta_{T_{k+1}}(\mathbf{Q}(t_k)) - V_g \mathbb{E} \left[ \sum_{\tau=0}^{T_{k+1}-1} g(\mathbf{r}(t_k + \tau)) \mid \mathbf{Q}(t_k) \right] \\ \leq B - f_{\text{QRRNUM}}(\mathbf{Q}(t_k)) - g_{\text{QRRNUM}}(\mathbf{Q}(t_k)) \\ \stackrel{(a)}{\leq} B - \mathbb{E} \left[ T_{k+1} \left( V_g g(\bar{\mathbf{y}}^*(\epsilon)) + \epsilon \sum_{n=1}^N Q_n(t_k) \right) \mid \mathbf{Q}(t_k) \right] \end{aligned} \quad (34)$$

where (a) is from (33). Taking expectation over  $\mathbf{Q}(t_k)$  in (34) and summing it over  $k \in \{0, \dots, K-1\}$ , we get

$$\begin{aligned} \mathbb{E} [L(\mathbf{Q}(t_K))] - \mathbb{E} [L(\mathbf{Q}(t_0))] - V_g \mathbb{E} \left[ \sum_{\tau=0}^{t_K-1} g(\mathbf{r}(\tau)) \right] \\ \leq BK - V_g g(\bar{\mathbf{y}}^*(\epsilon)) \mathbb{E} [t_K] - \epsilon \mathbb{E} \left[ \sum_{k=0}^{K-1} T_{k+1} \sum_{n=1}^N Q_n(t_k) \right]. \end{aligned} \quad (35)$$

Since  $Q_n(\cdot)$  and  $L(\mathbf{Q}(\cdot))$  are nonnegative and  $\mathbf{Q}(t_0) = \mathbf{0}$ , ignoring all backlog-related terms in (35) yields

$$\begin{aligned} -V_g \mathbb{E} \left[ \sum_{\tau=0}^{t_K-1} g(\mathbf{r}(\tau)) \right] & \leq BK - V_g g(\bar{\mathbf{y}}^*(\epsilon)) \mathbb{E} [t_K] \\ & \stackrel{(a)}{\leq} B \mathbb{E} [t_K] - V_g g(\bar{\mathbf{y}}^*(\epsilon)) \mathbb{E} [t_K] \end{aligned} \quad (36)$$

where (a) uses  $t_K = \sum_{k=1}^K T_k \geq K$ . Dividing (36) by  $V_g$  and rearranging terms, we get

$$\mathbb{E} \left[ \sum_{\tau=0}^{t_K-1} g(\mathbf{r}(\tau)) \right] \geq \left( g(\bar{\mathbf{y}}^*(\epsilon)) - \frac{B}{V_g} \right) \mathbb{E} [t_K]. \quad (37)$$

Recall from Section IV-A that  $B$  is an unspecified constant satisfying  $B \geq N \mathbb{E} [T_k^2 \mid \mathbf{Q}(t)]$ . From (26) it suffices to define  $B \triangleq N \mathbb{E} [T_{\max}^2]$ .

In QRRNUM, let  $K(t)$  denote the number of transmission rounds ending before time  $t$ . Using  $t_{K(t)} \leq t < t_{K(t)+1}$ , we have  $0 \leq t - \mathbb{E} [t_{K(t)}] \leq \mathbb{E} [t_{K(t)+1} - t_{K(t)}] = \mathbb{E} [T_{K(t)+1}]$ . Dividing the above by  $t$  and passing  $t \rightarrow \infty$ , we get

$$\lim_{t \rightarrow \infty} \frac{t - \mathbb{E} [t_{K(t)}]}{t} = 0. \quad (38)$$

<sup>7</sup>This property is proved in a similar case in [16, Ch. 5.5.2].

<sup>8</sup>We note that this renewal reward process is defined solely with respect to the service policy  $\text{RandRR}_\epsilon^*$ , and the network state needs not renew itself at the renewal epochs.

<sup>9</sup>Since the throughput vector  $\bar{\mathbf{y}}^*(\epsilon) = (\bar{y}_n^*(\epsilon))_{n=1}^N$  is achievable in  $\Lambda_{\text{int}}(\epsilon)$ , each component  $\bar{y}_n^*(\epsilon)$  must be less than or equal to the stationary probability  $\pi_{n,\text{ON}} \leq 1$ , and thus is a feasible choice of  $r_n(t)$ .

Next, the expected sum utility over the first  $t$  slots satisfies

$$\begin{aligned} \sum_{\tau=0}^{t-1} \mathbb{E}[g(\mathbf{r}(\tau))] &= \mathbb{E} \left[ \sum_{\tau=0}^{t_{K(t)}-1} g(\mathbf{r}(\tau)) \right] + \mathbb{E} \left[ \sum_{\tau=t_{K(t)}}^{t-1} g(\mathbf{r}(\tau)) \right] \\ &\stackrel{(a)}{\geq} \left[ g(\bar{\mathbf{y}}^*(\epsilon)) - \frac{B}{V_g} \right] \mathbb{E}[t_{K(t)}] \\ &= \left[ g(\bar{\mathbf{y}}^*(\epsilon)) - \frac{B}{V_g} \right] t - \left[ g(\bar{\mathbf{y}}^*(\epsilon)) - \frac{B}{V_g} \right] (t - \mathbb{E}[t_{K(t)}]), \end{aligned} \quad (39)$$

where (a) uses (37) and that  $g(\cdot)$  is nonnegative. Dividing (39) by  $t$ , taking a  $\liminf$  as  $t \rightarrow \infty$  and using (38), we get

$$\liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}[g(\mathbf{r}(\tau))] \geq g(\bar{\mathbf{y}}^*(\epsilon)) - \frac{B}{V_g}. \quad (40)$$

Using Jensen's inequality and the concavity of  $g(\cdot)$ , we get

$$\liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}[g(\mathbf{r}(\tau))] \leq \liminf_{t \rightarrow \infty} g(\bar{\mathbf{r}}^{(t)}), \quad (41)$$

where we define the average admission data vector:

$$\bar{\mathbf{r}}^{(t)} \triangleq (\bar{r}_n^{(t)})_{n=1}^N, \quad \bar{r}_n^{(t)} \triangleq \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}[\mathbf{r}(\tau)]. \quad (42)$$

Combining (40)(41) yields

$$\liminf_{t \rightarrow \infty} g(\bar{\mathbf{r}}^{(t)}) \geq g(\bar{\mathbf{y}}^*(\epsilon)) - \frac{B}{V_g},$$

which holds for any sufficiently small  $\epsilon$ . Passing  $\epsilon \rightarrow 0$  yields

$$\liminf_{t \rightarrow \infty} g(\bar{\mathbf{r}}^{(t)}) \geq g(\bar{\mathbf{y}}^*) - \frac{B}{V_g}. \quad (43)$$

Finally, we show the network is stable, and as a result

$$\liminf_{t \rightarrow \infty} g(\bar{\mathbf{y}}^{(t)}) \geq \liminf_{t \rightarrow \infty} g(\bar{\mathbf{r}}^{(t)}), \quad (44)$$

where  $\bar{\mathbf{y}}^{(t)} = (\bar{y}_n^{(t)})_{n=1}^N$  is defined similarly as  $\bar{\mathbf{r}}^{(t)}$ . Then combining (43)(44) finishes the proof. To prove stability, ignoring the first, second, and fifth term in (35) yields

$$\begin{aligned} \epsilon \mathbb{E} \left[ \sum_{k=0}^{K-1} T_{k+1} \sum_{n=1}^N Q_n(t_k) \right] &\leq BK + V_g \mathbb{E} \left[ \sum_{\tau=0}^{t_{K(t)}-1} g(\mathbf{r}(\tau)) \right] \\ &\stackrel{(a)}{\leq} K(B + V_g G_{\max} \mathbb{E}[T_{\max}]) \end{aligned} \quad (45)$$

where we define  $G_{\max} \triangleq g(\mathbf{1}) < \infty$  as the maximum value of  $g(\cdot)$  (since  $g(\cdot)$  is nondecreasing), and (a) uses

$$g(\mathbf{r}(\tau)) \leq G_{\max}, \quad \mathbb{E}[t_K] = \sum_{k=1}^K \mathbb{E}[T_k] \leq K \mathbb{E}[T_{\max}].$$

Dividing (45) by  $K\epsilon$ , taking a  $\limsup$  as  $K \rightarrow \infty$ , and using  $T_{k+1} \geq 1$ , we get

$$\limsup_{K \rightarrow \infty} \frac{1}{K} \mathbb{E} \left[ \sum_{k=0}^{K-1} \sum_{n=1}^N Q_n(t_k) \right] \leq \frac{B + V_g G_{\max} \mathbb{E}[T_{\max}]}{\epsilon} < \infty. \quad (46)$$

Equation (46) shows that the average backlog is bounded when sampled at time instants  $\{t_k\}$ . This property is enough to conclude that the average backlog over the whole time horizon is bounded, namely (8) holds and the network is stable. It is because the length of each transmission round  $T_k$  has a finite second moment and the maximum amount of data admitted to each user in every slot is at most 1; see [13, Lemma 13] for a detailed proof.

It remains to show network stability leads to (44). Recall that  $y_n(\tau) = \min[Q_n(\tau), \mu_n(\tau)]$  is the number of user- $n$  packets served in slot  $\tau$ , and (7) is equivalent to

$$Q_n(\tau+1) = Q_n(\tau) - y_n(\tau) + r_n(\tau). \quad (47)$$

Summing (47) over  $\tau \in \{0, \dots, t-1\}$ , taking an expectation and dividing it by  $t$ , we get

$$\frac{\mathbb{E}[Q_n(t)]}{t} = \bar{r}_n^{(t)} - \bar{y}_n^{(t)}, \quad (48)$$

where  $\bar{r}_n^{(t)}$  is defined in (42) and  $\bar{y}_n^{(t)}$  is defined similarly. From [25, Theorem 4(c)], the stability of  $Q_n(t)$  and (48) result in that for each  $n$ :

$$\limsup_{t \rightarrow \infty} \frac{\mathbb{E}[Q_n(t)]}{t} = \limsup_{t \rightarrow \infty} (\bar{r}_n^{(t)} - \bar{y}_n^{(t)}) = 0. \quad (49)$$

Noting that  $g(\cdot)$  is bounded, there exists a convergent subsequence of  $g(\bar{\mathbf{y}}^{(t)})$  indexed by  $\{t_i\}_{i=1}^\infty$  such that

$$\lim_{i \rightarrow \infty} g(\bar{\mathbf{y}}^{(t_i)}) = \liminf_{t \rightarrow \infty} g(\bar{\mathbf{y}}^{(t)}). \quad (50)$$

By iteratively finding a convergent subsequence of  $\{\bar{r}_n^{(t_i)}\}_{i=1}^\infty$  for each  $n$  (noting that  $\bar{r}_n^{(t)}$  is bounded for all  $n$  and  $t$ ), there exists a subsequence  $\{t_k\} \subset \{t_i\}$  such that  $\{\bar{\mathbf{r}}^{(t_k)}\}_{k=1}^\infty$  converges as  $k \rightarrow \infty$ . From (49) and that  $\limsup\{z_n\}$  is the supremum of all limit points of a sequence  $\{z_n\}$ , we get

$$\lim_{k \rightarrow \infty} (\bar{r}_n^{(t_k)} - \bar{y}_n^{(t_k)}) \leq 0 \Rightarrow \lim_{k \rightarrow \infty} \bar{r}_n^{(t_k)} \leq \lim_{k \rightarrow \infty} \bar{y}_n^{(t_k)}, \quad \forall n. \quad (51)$$

It follows that

$$\begin{aligned} \liminf_{t \rightarrow \infty} g(\bar{\mathbf{y}}^{(t)}) &\stackrel{(a)}{=} \lim_{k \rightarrow \infty} g(\bar{\mathbf{y}}^{(t_k)}) \\ &\stackrel{(b)}{\geq} \lim_{k \rightarrow \infty} g(\bar{\mathbf{r}}^{(t_k)}) \stackrel{(c)}{\geq} \liminf_{t \rightarrow \infty} g(\bar{\mathbf{r}}^{(t)}), \end{aligned}$$

where (a) is from (50), (b) uses (51) and that  $g(\cdot)$  is continuous and nondecreasing, and (c) uses that  $\liminf\{z_n\}$  is the infimum of limit points of  $\{z_n\}$ . ■

## VI. CONCLUSIONS

We have provided a theoretical framework to do network utility maximization over partially observable Markov ON/OFF channels. The performance and control decisions in such networks are constrained by the limiting channel probing capability and delayed/uncertain channel state information, but can be improved by taking advantage of channel memory. Overall, to attack such problems we need to solve (at least approximately) high-dimensional restless bandit problems with a general functional objective, which are difficult to analyze using existing tools such as Whittle's index theory or Markov



decision theory. In this paper we propose a new methodology to solve such problems by combining an achievable region approach from mathematical programming and the powerful Lyapunov optimization theory. The key idea is to first identify a good constrained performance region rendered by stationary policies, and then solve the problem only over the constrained region, serving as an approximation to the original problem. While a constrained performance region is constructed in [13], in this paper using a novel frame-based variable-length Lyapunov drift argument, we can solve the original problem over the constrained region by constructing queue-dependent greedy algorithms that stabilize the network with near-optimal utility. It will be interesting to see how the Lyapunov optimization theory can be extended and used to attack other sequential decision making problems as well as stochastic network optimization problems with limited channel probing and delayed/uncertain channel state information.

## REFERENCES

- [1] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 396–409, Apr. 2008.
- [2] C.-P. Li and M. J. Neely, "Energy-optimal scheduling with dynamic channel acquisition in wireless downlinks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 4, pp. 527–539, Apr. 2010.
- [3] P. Chaporkar, A. Proutiere, H. Asnani, and A. Karandikar, "Scheduling with limited information in wireless systems," in *ACM Int. Symp. Mobile Ad Hoc Networking and Computing (MobiHoc)*, New Orleans, LA, May 2009.
- [4] A. Pantelidou, A. Ephremides, and A. L. Tits, "Joint scheduling and routing for ad-hoc networks under channel state uncertainty," in *IEEE Int. Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Apr. 2007.
- [5] L. Ying and S. Shakkottai, "On throughput optimality with delayed network-state information," in *Information Theory and Application Workshop (ITA)*, 2008, pp. 339–344.
- [6] —, "Scheduling in mobile ad hoc networks with topology and channel-state uncertainty," in *IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009.
- [7] M. Zorzi, R. R. Rao, and L. B. Milstein, "A Markov model for block errors on fading channels," in *Personal, Indoor and Mobile Radio Communications Symp. PIMRC*, Oct. 1996.
- [8] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probab.*, vol. 25, pp. 287–298, 1988.
- [9] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.
- [10] Q. Zhao and A. Swami, "A decision-theoretic framework for opportunistic spectrum access," *IEEE Wireless Commun. Mag.*, vol. 14, no. 4, pp. 14–20, Aug. 2007.
- [11] J. L. Ny, M. Dahleh, and E. Feron, "Multi-uav dynamic routing with partial observations using restless bandit allocation indices," in *American Control Conference*, Seattle, WA, USA, Jun. 2008.
- [12] C.-P. Li and M. J. Neely, "Exploiting channel memory for multi-user wireless scheduling without channel measurement: Capacity regions and algorithms," in *IEEE Int. Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Avignon, France, May 2010.
- [13] —, "Exploiting channel memory for multi-user wireless scheduling without channel measurement: Capacity regions and algorithms," arXiv report, 2010. [Online]. Available: <http://arxiv.org/abs/1003.2675>
- [14] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.
- [15] —, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 466–478, Mar. 1993.
- [16] M. J. Neely, "Dynamic power allocation and routing for satellite and wireless networks with time varying channels," Ph.D. dissertation, Massachusetts Institute of Technology, November 2003.
- [17] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends in Networking*, vol. 1, no. 1, 2006.
- [18] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of whittle's index for dynamic multichannel access," Tech. Rep., 2008.
- [19] J. Nino-Mora, "An index policy for dynamic fading-channel allocation to heterogeneous mobile users with partial observations," in *Next Generation Internet Networks, 2008. NGI 2008*, 2008, pp. 231–238.
- [20] S. Guha, K. Munagala, and P. Shi, "Approximation algorithms for restless bandit problems," Tech. Rep., Feb. 2009.
- [21] J. Nino-Mora, "Dynamic priority allocation via restless bandit marginal productivity indices," *TOP*, vol. 15, no. 2, pp. 161–198, 2007.
- [22] H. Yu, "Approximate solution methods for partially observable markov and semi-markov decision processes," Ph.D. dissertation, Massachusetts Institute of Technology, Feb. 2006.
- [23] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 2005, vol. I.
- [24] S. H. A. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4040–4050, Sep. 2009.
- [25] M. J. Neely, "Stability and capacity regions for discrete time queueing networks," arXiv report, Mar. 2010. [Online]. Available: <http://arxiv.org/abs/1003.3396>